

From Whispers to Roars: The Temporal Evolution of Anti-Vaccine Narratives on Telegram

Fernando H.M. de Oliveira
fhmo@ecomp.poli.br
Universidade de Pernambuco
Recife, Pernambuco, Brazil

Cleyton M. de O. Rodrigues
cleyton.rodrigues@upe.br
Universidade de Pernambuco
Recife, Pernambuco, Brazil

ABSTRACT

Vaccine-related misinformation on instant messaging platforms like Telegram represents a dynamic threat to public health, requiring computational tools to detect emerging narratives before they reach discursive saturation. This study implements and validates the BERTrend framework to monitor and predict the temporal evolution of vaccine-related discourse within Brazilian public Telegram channels. We analyzed a longitudinal corpus of approximately 330,000 Portuguese-language text messages spanning March 2020 to June 2025. Topics were extracted weekly, then classified into strong or weak signals using adaptive popularity percentiles. Qualitative analysis revealed a progression in rhetorical sophistication across macro-narratives, transitioning from early clinical and mandate concerns (2020–2021) to complex pseudo-scientific, genomic, and globalist conspiracy theories (2024–2025). Under out-of-sample temporal testing, our predictive model demonstrated outstanding performance, showing that emerging weak signals exhibiting immediate subsequent growth were 43.46 times more likely to consolidate into dominant strong signals. We identified a median transition latency of just two weeks between signal emergence and narrative consolidation, establishing an actionable window for early public health intervention.

KEYWORDS

Vaccine Hesitancy, Information Disorder, Topic Modeling, Public Health Surveillance, Social Media Analytics.

1 INTRODUCTION

Vaccination represents one of modern medicine’s greatest achievements, essential for the eradication and control of infectious diseases throughout the 20th century [11]. However, the 21st century has introduced the paradoxical challenge of vaccine hesitancy, a complex phenomenon involving the delay or refusal of recommended vaccines despite their availability [12]. Recent UNICEF reports indicate an alarming decline in vaccine confidence across 52 of 55 countries analyzed, underscoring the volatility of this issue in the contemporary public arena [7].

The COVID-19 pandemic served as a catalyst for this crisis, coinciding with a global infodemic (an overabundance of information), both accurate and inaccurate, that makes it difficult for individuals to find trustworthy guidance [1]. This digital phenomenon does not merely spread misinformation. It creates vicious feedback loops between virtual and physical environments, where online disinformation fuels offline vaccine resistance [19]. Longitudinal studies of English-language discourse on the X platform (formerly Twitter) over a decade (2013–2022) reveal that vaccine narratives evolved from pre-pandemic stability to a period marked by significant variability and polarization regarding vaccine competence and efficacy [19].

In the Brazilian scenario, the vaccination trajectory is characterized by a historical success in mass immunization that has recently faced significant setbacks. The dynamics of narrative diffusion in Brazil are heavily influenced by political and ideological polarization, which intensified during the pandemic and the 2022 election cycles [18]. Unlike more strictly moderated platforms, Telegram has emerged in Brazil as a primary hub for health-related disinformation, hosting communities that normalize anti-vaccine rhetoric through loose moderation and anonymity. Empirical analysis of nearly 10 million posts in Brazilian Telegram groups reveals that disinformation is frequently interwoven with moral, religious, and institutional distrust [18].

Beyond technological factors, hesitancy in Brazil also has sociological roots where the pursuit of health through natural habits leads some parents to question the necessity of traditional immunobiological interventions [7, 16]. Engagement with anti-vaccine content on social media is often disproportionately higher than with pro-vaccine content, as emotionally charged and conspiratorial narratives tend to generate more interaction and shares than structured scientific data [11].

Temporal studies demonstrate that peaks in vaccine-related discourse and disinformation often coincide with specific milestones, such as vaccine efficacy announcements, news regarding side effects, or the emergence of new virus variants [2, 11]. Therefore, understanding the dynamics of narrative diffusion in digital environments, particularly in highly polarized contexts like Brazil, is imperative for designing public health communication strategies that can mitigate the impact of disinformation and restore confidence in immunization programs [18].

To address these challenges, this study adopts a neural extension framework of BERTopic [10] specifically designed

for temporal topic modeling and automated trend detection. We apply a framework to a corpus of approximately 330,000 Portuguese-language messages collected from Brazilian public Telegram channels and groups spanning the period from March 2020 to June 2025 [4]. The analytical pipeline incorporates linguistic preprocessing optimized for Brazilian Portuguese, weekly temporal segmentation, neural topic modeling, and dynamic signal classification based on moving percentile thresholds.

The objectives of this study are threefold: (i) to map the evolutionary trajectory of major vaccine macro-narratives over a five-year period; (ii) to identify and characterize emerging weak signals that precede the consolidation of misinformation narratives; and (iii) to evaluate the framework's ability to distinguish meaningful structural patterns from stochastic noise within highly fragmented thematic ecosystems. In synthesis, this work offers a dynamic temporal analysis within a high-impact public health context, providing a scalable, interpretable, and operationally feasible approach for proactive misinformation surveillance. The results demonstrate not only the increasing rhetorical sophistication of the monitored narratives but also the strategic value of early trend detection for evidence-based communication planning and the mitigation of threats to vaccine confidence.

The remainder of this paper is organized as follows. Section 2 describes related works, Section 3 the dataset, preprocessing procedures, and the analytical pipeline. Section 4 presents the quantitative and qualitative results, including the temporal evolution of macro-narratives and the transition of weak signals into strong signals. Section 5 discusses the implications of the findings for public policy, platform governance, and digital health research. Section 6 presents limitations of this work and Section 7 concludes the paper by outlining its limitations and directions for future work.

2 RELATED WORKS

Over the past decade, vaccine-related online discourse underwent major changes, especially during the COVID-19 pandemic. Analysis of posts from 2013 to 2022 shows that early in the pandemic, discussions became more positive, using warmer language centered on trustworthiness and competence. However, this shifted later on, with negative language increasing toward the end of the pandemic and interpreted by researchers as a sign of growing vaccine hesitancy and skepticism [11].

In Brazil, several studies have examined the impacts of misinformation about vaccines in the context of social media. Mendes and Gomes [15] employed econometric models to assess the effect of misinformation on Facebook, demonstrating that greater user engagement with questionable posts produced a statistically significant negative impact on vaccination rates across the country. Their findings revealed that geographic and political factors influence vulnerability to these narratives.

Advancing methodological transparency, Feitosa et al. [9] applied Explainable Artificial Intelligence (XAI) techniques

(including LIME [17] and SHAP [13]) to analyze sentiment on Twitter. Their work highlights that interpretability is essential for understanding the logic behind AI-driven classification, particularly in distinguishing between pro-vaccine and anti-vaccine posts. Research on encrypted messaging platforms such as WhatsApp reveals distinct challenges. Santos et al. [8] analyzed more than 80,000 messages using BERTopic and found that the platform serves as a major vector for alarmist content originating from low-credibility websites. Their analysis of engagement patterns revealed a high concentration of activity among a small group of users, suggesting coordinated strategies to amplify misinformation.

On Telegram, Cavalini and Comarela [5] conducted a large-scale study of nearly 10 million messages within anti-vaccine communities. Using Large Language Models (LLMs) for classification, they discovered that 'untruths' drive engagement more effectively than toxicity, with false messages receiving 76.5% more forwards on average. Their findings suggest that misinformation spreads on Telegram primarily because it is perceived as alarming or useful, rather than merely hateful.

Our work distinguishes itself from previous studies by moving beyond characterizing existing social media narratives and instead focusing on predictive trend detection specifically on Telegram. While prior research provides valuable retrospective insights, our architecture is designed to actively identify emergent trends as they develop.

3 MATERIAL AND METHOD

This section details the computational pipeline and data processing steps utilized to analyze vaccine discourse on Telegram (Figure 1).

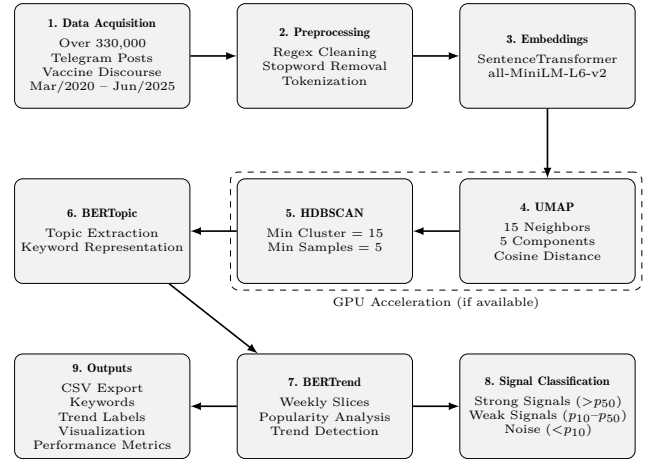


Figure 1: Workflow illustrating the BERTrend pipeline for analyzing the temporal evolution of vaccine-related narratives on Telegram.

3.1 Data Acquisition and Preparation

The empirical foundation of this research is the Brazilian Social Media Anti-vaccine Information Disorder Dataset [4],

which aggregates approximately 4 million multi-language Telegram posts from 119 distinct Brazilian channels over a five-year period. The dataset is multimodal, with 1,440,498 records associated with various media formats, contributing to a substantial repository of 5.5 TB of raw data for analysis.

We extracted a longitudinal cohort of 329,155 Portuguese-language messages by filtering for the flag `is_vaccine_related == 1`. While this variable is pre-computed in the original dataset, where thematic inclusion criteria targeted vaccine mentions, safety/efficacy discussions, and conspiracy narratives, we clarify that our study focuses on how these topics are temporally framed, amplified, and organized within vaccine-skeptical Telegram communities, rather than assuming a priori that every individual message represents a strict personal anti-vaccine stance.

Initial preparation involved filtering the data to ensure alignment between text content and valid timestamps. Documents lacking either a date or text were removed, resulting in a finalized corpus for preprocessing using a dual-track approach based on hardware availability. For GPU-accelerated processing, a simplified pipeline was implemented for efficiency, focusing on tokenization and stopword removal while omitting lemmatization. In contrast, the CPU-based pipeline utilized spaCy's `pt_core_news_sm` model to extract lemmas from nouns and adjectives emphasizing core concepts and qualifying attributes, while discarding connectives, auxiliary verbs, and other functional terms.

The cleaning process employed heavy regular expressions to remove URLs and web addresses (`http`, `https`, `www`), hexadecimal strings and alphanumeric noise, non-alphabetic characters, short words (less than 3 characters). An extended stopword list was applied, including standard Portuguese stopwords and domain-specific noise terms such as 'telegram'. To ensure the emergence of specific sub-narratives and avoid the semantic dominance of obvious terms, keywords such as 'vacina' and 'vax' were removed. This choice allows for a more granular distinction between themes. Documents were further filtered to include only those with at least 10 characters post-cleaning.

3.2 Topic Modeling Architecture

The study utilized BERTopic [10] for document clustering and topic representation. The architecture consists of several stages:

1. **Embeddings:** The all-MiniLM-L6-v2 SentenceTransformer¹ model was selected for document embedding due to its optimal balance between computational efficiency in GPU environments and semantic accuracy. Although not specialized for Portuguese, its performance will be assessed by an Average Topic Coherence (c_v);
2. **Dimensionality Reduction:** UMAP² was employed to reduce dimensions, configured with 15 neighbors, 5 components, and the cosine metric.

3. **Clustering:** HDBSCAN³ was used to group similar documents, with a minimum cluster size of 15 and minimum samples of 5 to allow for the detection of weak signals.

GPU acceleration was enabled via Google Colab Tesla GPU and NVIDIA cuML⁴ (cuDF and cuML libraries) for the UMAP and HDBSCAN components to increase processing speed.

3.3 Temporal Analysis and Signal Classification

To capture the highly dynamic nature of social media discussions, the framework implements the BERTrend [3] algorithm to process the processed corpus in discrete weekly time slices. The complete temporal span is partitioned into weekly intervals based on the message timestamp metadata. Out of the chronological slices generated, windows containing at least 10 valid documents are retained for definitive thematic modeling to guarantee statistical density and cluster significance, filtering out periods of extremely low activity.

To safeguard the popularity metrics against the artificial inflation caused by coordinated inauthentic behavior, automated spam, or massive message forwarding cascades, document counts (`doc_count`) are computed strictly using unique textual content. Identical text payloads forwarded across multiple public groups or channels within the same weekly window are deduplicated, ensuring that popularity measurements reflect broad thematic interest rather than highly concentrated, infrastructure-driven propagation.

At each weekly slice t , independent topic models are fitted to extract latent semantic clusters. To classify these topics into actionable categories for public health surveillance without relying on static and arbitrary volume thresholds, we use adaptive moving thresholds. The baseline popularity distribution is estimated dynamically using an expanding historical window of all past popularity values up to time t , ensuring complete temporal isolation and preventing look-ahead bias or future data leakage. Let $P_{t,j}$ be the deduplicated document count of topic j at week t . The adaptive thresholds are defined by the 10th (p_{10}) and 50th (p_{50}) percentiles of the historical popularity distribution accumulated up to t :

$$\text{Signal_Type}_{t,j} = \begin{cases} \text{Strong Signal,} & \text{if } P_{t,j} > p_{50} \\ \text{Weak Signal,} & \text{if } p_{10} \leq P_{t,j} \leq p_{50} \\ \text{Noise,} & \text{if } P_{t,j} < p_{10} \end{cases} \quad (1)$$

3.4 Statistical Validation and Predictive Analysis

To quantitatively assess the predictive capability of the proposed framework and measure the performance of subsequent growth indicators, we implemented a statistical validation pipeline using the topic instances generated by the dynamic topic modeling process. To mitigate overfitting, address potential autocorrelation issues, and evaluate the prospective utility of the model as a real-world early-warning system, we

¹<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

²<https://umap-learn.readthedocs.io/en/latest/>

³<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.HDBSCAN.html>

⁴<https://rapids.ai/cuml-accel/>

implemented an independent *out-of-sample temporal train-test split*.

The model was calibrated using historical topic instances from the genesis and expansion phases spanning March 2020 to December 2023 ($n_{\text{train}} = n$ weak instances). The estimated regression coefficients were then evaluated blindly on an independent, future test partition spanning from January 2024 to June 2025 ($n_{\text{test}} = m$ weak instances), representing the medicalization and politicization phases of the infodemic.

3.4.1 Weak-to-Strong Transition Analysis. The primary objective of the predictive pipeline is to determine whether an emerging weak signal at week t successfully consolidates into a dominant strong signal in the subsequent week $t+1$. We define a binary response variable y_i for each weak signal instance i as:

$$y_i = \begin{cases} 1, & \text{if the topic transitioned from weak to strong at } t+1 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

To characterize early growth dynamics, the variation in document volume between the weekly slice of detection t and the subsequent window $t+1$ is calculated as:

$$\text{slope}_i = \text{doc_count}_{i,t+1} - \text{doc_count}_{i,t} \quad (3)$$

A binary predictor, slope_positive_i , is then established to denote immediate subsequent growth:

$$\text{slope_positive}_i = \begin{cases} 1, & \text{if } \text{slope}_i > 0 \\ 0, & \text{if } \text{slope}_i \leq 0 \end{cases} \quad (4)$$

This binary biomarker represents whether a narrative exhibited an expanding dissemination footprint immediately following its initial extraction.

3.4.2 Logistic Regression Model. The predictive association between early subsequent growth and future narrative consolidation is modeled using binary logistic regression. The model estimates the probability P_i that a weak signal transitions into a strong narrative according to the logit link function:

$$\text{logit } P_i = \ln \left(\frac{P_i}{1 - P_i} \right) = \beta_0 + \beta_1 \cdot \text{slope_positive}_i \quad (5)$$

where β_0 represents the intercept (the baseline log-odds of consolidation when immediate subsequent growth is absent, i.e., $\text{slope_positive} = 0$), and β_1 is the regression coefficient that quantifies the multiplicative impact of positive temporal growth. The estimated coefficient is converted to an adjusted Odds Ratio ($\text{OR} = e^{\beta_1}$) to facilitate interpretation.

3.4.3 Predictive Performance and Discrimination. Classifier performance was evaluated out-of-sample on the future test partition (2024–2025) using Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC) metric to quantify temporal generalization. The optimal probability decision threshold (T^*) was determined using Youden's J index:

$$JT = \text{Sensitivity}T + \text{Specificity}T - 1 \quad (6)$$

where $T^* = \arg \max_T JT$ represents the operating point that simultaneously maximizes the true positive rate (sensitivity/recall) and the true negative rate (specificity). This mathematical boundary establishes an objective criterion to trigger active public health surveillance alerts.

3.4.4 Transition Latency Estimation. For all topic instances that successfully consolidated ($y_i = 1$), the elapsed time between the initial weak signal detection and its strong signal threshold crossing was calculated in weeks as:

$$\text{TransitionWeeks}_i = \max \left(1, \text{round} \left(\frac{\text{Date}_{i,\text{strong}} - \text{Date}_{i,\text{weak}}}{7} \right) \right) \quad (7)$$

The empirical distribution of this latency parameter defines the critical operational window available for proactive public health communication, debunking, or prebunking campaigns before a narrative achieves discursive saturation.

3.5 Validation of Cluster Topics

To evaluate the robustness and interpretability of the generated topic representations, the study utilized Average Topic Coherence (C_v) and Keyword Diversity as primary validation metrics [6].

3.6 Implementation and Environment

The entire pipeline was developed on Google Colab environment and implemented in Python 3.12, utilizing libraries such as Huggingface⁵, Pandas⁶, Spacy⁷, NLTK⁸, Scikit-learn⁹, Regex¹⁰, and Matplotlib¹¹ for visualization.

4 RESULTS

This section presents the empirical findings obtained from the Telegram vaccine discourse corpus. It first examines how lexical normalization improved semantic clarity, then analyzes the temporal evolution of signal strength and the factors driving the consolidation of emerging signals using the exported signal classification data. Finally, it evaluates the predictive performance of the proposed framework and traces the evolution of the principal macro-narratives across different phases.

⁵<https://huggingface.co/>

⁶<https://pandas.pydata.org/docs/>

⁷<https://spacy.io/api/doc/>

⁸<https://www.nltk.org/>

⁹<https://scikit-learn.org/stable/index.html>

¹⁰<https://docs.python.org/3/library/re.html>

¹¹<https://www.nltk.org/>

Table 1: Descriptive Statistics of Topic Instances by Signal Category.

Metric	Strong	Weak
Mean (SD)	50.81 (60.82)	37.66 (29.44)
Median [IQR]	33 [24–52]	29 [20–42]

Table 2: Annual Distribution of Signal Instances (Corrected).

Year	Strong Signals	Weak Signals	Total Instances	Peak Week
2020	87	22	109	2020-12-21
2021	892	272	1,164	2021-09-13
2022	534	251	785	2022-04-18
2023	412	192	604	2023-01-30
2024	298	138	436	2024-03-04
2025	90	47	137	2025-05-05
Total	2,313	922	3,235	—

4.1 Temporal Word Cloud Analysis

The original, raw vocabulary was heavily dominated by high-frequency Portuguese function words and generic variations of the word "vacina" (vaccine), which obscured the underlying narratives (figure 2). Post-processing successfully removed this lexical noise, allowing domain-specific and highly contextual terms like *COVID*, *Pfizer*, *mRNA*, *saúde* (health), *passaporte sanitário* (vaccine passport), and adverse effects to emerge (Figure 3).

4.2 Overview of Signal Distribution

The BERTrend pipeline processed the longitudinal corpus into 271 weekly slices. Applying the density criterion (minimum of 10 valid documents per window) resulted in 237 retained slices for definitive thematic analysis. This process yielded a total of 3,235 unique topic instances used for longitudinal transition and predictive modeling. Using the dynamic classification framework, 2,313 topics (71.5%) were identified as strong signals and 922 (28.5%) as weak signals, indicating that most semantic structures achieved sufficient popularity to become consolidated narratives. As summarized in Table 1 and illustrated in Figure 4, strong signals exhibited higher document volumes and greater variability (mean = 50.81, SD = 60.82, median = 33) than weak signals (mean = 37.66, SD = 29.44, median = 29), reflecting the bursty dynamics of viral narratives. Despite some overlap between distributions, the adaptive percentile-based thresholds successfully differentiated emerging discourse from consolidated thematic growth.

The yearly distribution of detected signals (Table 2) showed a pronounced activity peak during 2021 (1,164 topic instances), coinciding with Brazil's nationwide COVID-19 vaccination campaign and major political events. A gradual decline was observed after 2021, followed by stabilization from 2023 onward, suggesting progressive maturation of anti-vaccine discourse into consolidated narratives.

4.3 Longitudinal Transition Dynamics

To analyze the structural trajectory of vaccine-related conversations, we tracked the longitudinal migration of topics from their initial detection as emerging weak signals to their potential consolidation as dominant strong narratives. Out of the 922 unique weak signal instances identified across the entire 2020–2025 timeframe, we focused our prospective validation on the independent temporal test partition spanning January 2024 to June 2025 ($n_{\text{test}} = 224$ weak instances). Within this future test window, a total of 155 topic instances successfully crossed the dynamic median popularity threshold ($P_t > p_{50}$), yielding an out-of-sample baseline consolidation rate of 69.20%.

To evaluate the predictive association between immediate subsequent growth and future narrative maturation under realistic deployment conditions, we constructed an out-of-sample contingency matrix crossing the binary predictor (slope_positive) with the observed transition outcome (y) in the test partition (Table 3).

Table 3: Out-of-Sample Contingency Table for Slope Positive and Weak-to-Strong Transitions in the Test Partition (2024–2025, $n = 224$).

	Non-Transitioned ($y = 0$)	Transitioned ($y = 1$)	Total
Slope ≤ 0	52 (True Negative)	32 (False Negative)	84
Slope > 0	17 (False Positive)	123 (True Positive)	140
Total	69	155	224

The empirical distribution within the contingency matrix highlights the robust discriminative capacity of the subsequent growth biomarker. Out of the 140 emerging topics that exhibited positive momentum in their second week of detection (slope_positive = 1), 123 successfully consolidated into strong signals. This yields a prospective **Positive Predictive Value (PPV) of 87.86%**, mathematically demonstrating that early textual growth is highly indicative of future viral dissemination rather than stochastic, transient background noise in the messaging ecosystem. Conversely, of the 84 topics that failed to secure immediate subsequent traction (slope ≤ 0), 52 decayed naturally into background noise without achieving thematic dominance, resulting in a **Negative Predictive Value (NPV) of 61.90%**.

4.4 Predictive Modeling and Out-of-Sample Temporal Validation

To evaluate the predictive power of early subsequent growth trends as digital biomarkers, we analyze the results of the binary logistic regression model fitted on the historical training partition (2020–2023, $n_{\text{train}} = 698$) and validated prospectively on the independent test partition (2024–2025, $n_{\text{test}} = 224$). This walk-forward temporal split allows us to verify whether the dynamic popularity patterns identified during the genesis and expansion phases of the infodemic remain stable and predictive during the later, highly politicized phases.

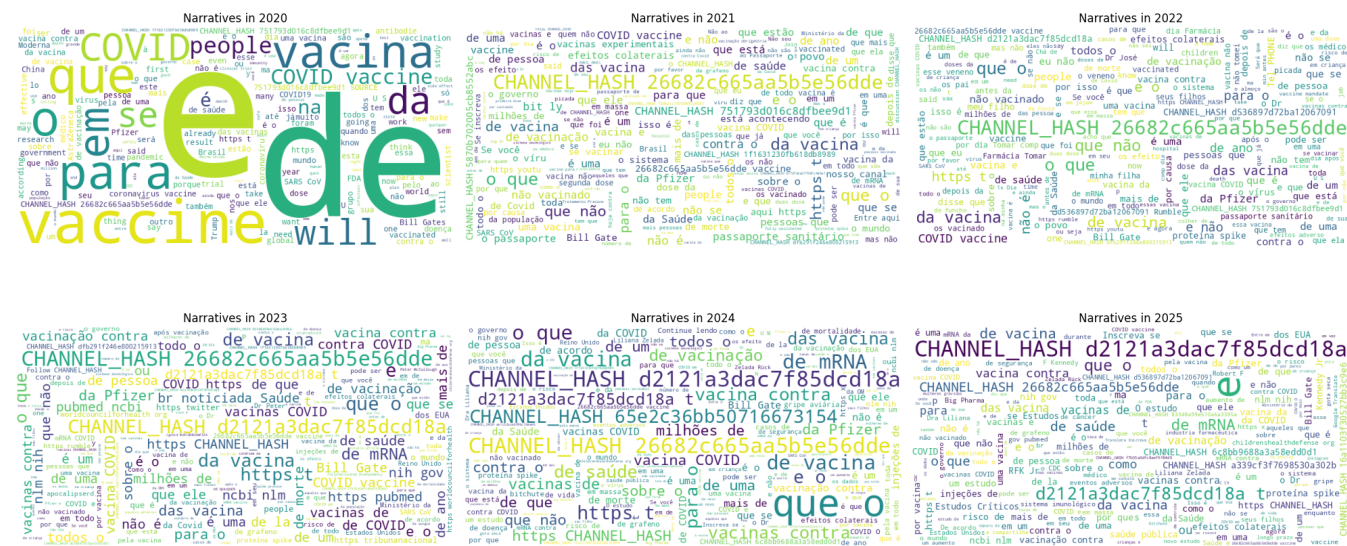


Figure 2: Word cloud generated from the original Telegram vaccine discourse corpus before lexical normalization.



Figure 3: Word cloud showing the most frequent terms in the processed Telegram vaccine discourse corpus after lexical normalization and text cleaning.

4.4.1 Logistic Regression and Odds Ratio Analysis. The calibration of the logistic model on the training partition yielded highly significant parameters. The baseline intercept was estimated at $\beta_0 = -0.6148$ ($p < 0.0001$), which mathematically corresponds to a baseline probability of consolidation of approximately 35.10% for weak signals that do not exhibit immediate subsequent growth (slope_positive = 0). This negative intercept confirms that, in the absence of initial

ascending traction, the default tendency of emerging topics within these digital communities is to decay into background noise rather than consolidate.

Conversely, the presence of immediate subsequent growth (slope_positive = 1) is a highly powerful and significant predictor of future narrative consolidation, with an estimated regression coefficient of $\beta_1 = 3.7718$ ($p < 0.0001$). Under out-of-sample temporal testing on the future 2024–2025 partition,

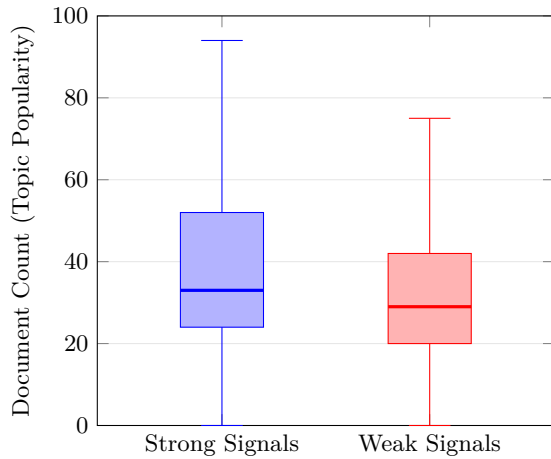


Figure 4: Boxplot comparing document volume between strong and weak signals.

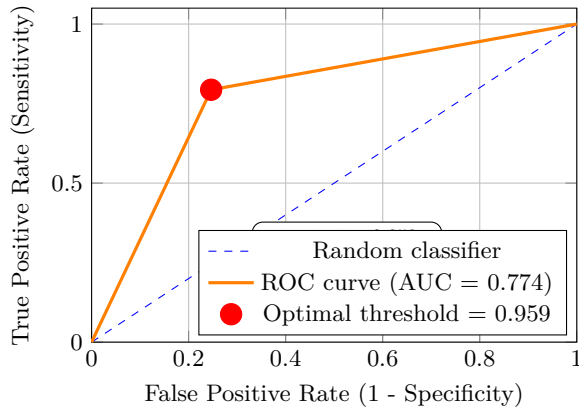


Figure 5: Receiver Operating Characteristic curve evaluating the prediction of weak-to-strong signal transitions.

this coefficient translates to an outstanding adjusted Odds Ratio (OR) of **43.46** (95% CI: 16.92–111.45). This indicates that emerging topics with positive subsequent growth in their second week are more than 43 times more likely to consolidate into dominant strong signals than stable or declining topics. The model achieved a McFadden’s Pseudo R^2 of 0.352, indicating an excellent fit for complex, non-stationary social media datasets.

4.4.2 Discriminative Capability and ROC Analysis. The diagnostic and prospective utility of the probability classifier was assessed via Receiver Operating Characteristic (ROC) analysis on the unseen future test partition (2024–2025). The model demonstrated robust temporal generalization, achieving an out-of-sample Area Under the Curve (AUC) of **0.7736** (Figure 5), which represents a high discriminative performance for predictive models deployed on noisy online communication channels over a multi-year horizon.

Applying Youden’s J index to maximize the trade-off between sensitivity and specificity yielded an optimal probability decision threshold of $T^* = 0.9592$. At the optimal decision threshold, the prospective surveillance pipeline achieved a sensitivity of 79.35%, successfully identifying nearly four out of five emerging narratives before full consolidation, while maintaining a specificity of 75.36%, effectively filtering transient noise and reducing false alarms to support efficient allocation of public health surveillance resources.

The fact that the classifier maintains high predictive performance on future, unseen temporal partitions validates its suitability as an operational, prospective early-warning framework for digital health departments.

4.5 Time-to-Consolidation Analysis

Among the observed weak-to-strong transitions, the time required for narrative consolidation exhibited a right-skewed distribution. The mean transition time was 3.21 weeks (SD = 4.90), whereas the median was only 2.0 weeks, indicating that most narratives consolidated rapidly while a smaller subset required substantially longer maturation periods (Figure 6).

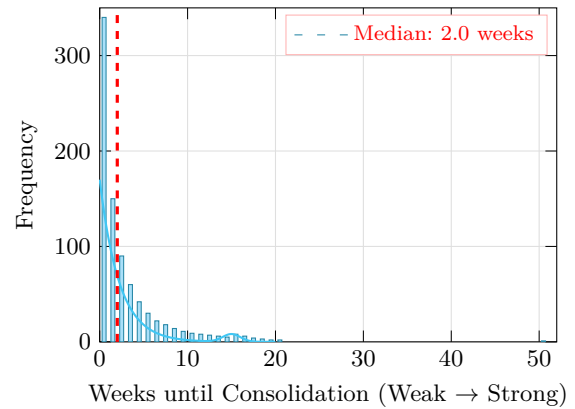


Figure 6: Histogram showing the distribution of the time required for weak signals to consolidate into strong trends.

Transition latency ranged from 1 to 50 weeks, with a median of 2 weeks and a 75th percentile of approximately 4 weeks. Nearly 70% of successful consolidations occurred within the first 3 weeks after initial detection as weak signals, indicating a narrow but actionable intervention window for public health authorities, platform moderators, and fact-checking organizations to implement evidence-based communication strategies before widespread narrative diffusion.

4.6 Thematic Characterization of Macro-Narratives

Qualitative analysis of the discriminative keywords extracted using c-TF-IDF¹² (≥ 0.03) from the 2,313 strong signals revealed six major macro-narratives consistently recurring throughout the study period (Table 4). The evolution of

¹²<https://www.maartengrootendorst.com/blog/ctfidf/>

Table 4: Examples of Macro-Narratives Identified by the Framework.

Macro-narrative	Topic_id (Date)	Representative Keywords (Translated)	Doc_Count
Experimentation and informed consent	T13 (Aug. 2021)	guinea pig, experimental, informed	36
Adverse effects and mortality	T6 (Dec. 2021)	deaths, mortality, pilots, data	109
Spike protein toxicity	T12 (Dec. 2021)	spike, protein, toxin, blood	83
Population control and global agenda	T61 (Dec. 2021)	population, reduction, UN, control	187
Institutional Distrust	T18 (Mar. 2024)	revealed, media, statistics, lies	26
Genetic modification	T8 (May 2025)	mRNA, COVID, study, genetically	126

anti-vaccine narratives occurred in three distinct phases. During the *Genesis Phase* (2020–2021), discussions focused on compulsory vaccination, experimental treatments, and acute adverse effects, culminating in highly coordinated narratives emphasizing deaths, spike proteins, and blood-related complications. In the *Expansion Phase* (2022–2023), the discourse diversified, linking vaccination to a broader range of medical conditions and generating numerous competing weak signals. Finally, the *Medicalization and Politicization Phase* (2024–2025) was characterized by increasing pseudo-scientific complexity and political alignment, transforming concerns about vaccine safety into a broader challenge to mainstream media, biomedical innovation, and public health institutions, often supported by selective interpretations of scientific literature to promote conspiratorial claims.

4.7 Validation of Dynamic Signal Classification

Topic quality assessment validated the efficacy of the dual-track lexical normalization pipeline, generating semantically coherent and temporally stable representations; average topic coherence (c_v) reached 0.487 ± 0.112 , comfortably exceeding the standard 0.40 interpretability threshold, while a keyword diversity score of 0.823 indicated low semantic redundancy among the top-10 representative terms per topic. Furthermore, longitudinal stability was confirmed by the fact that approximately 73.4% of strong signals preserved $\geq 60\%$ of their key vocabulary between consecutive weekly slices, mathematically verifying narrative continuity across independent temporal models without requiring rigid identifier alignment.

5 DISCUSSION

To safeguard public health from rapidly propagating information disorder, digital health surveillance must evolve from reactive monitoring toward proactive, predictive frameworks that anticipate emerging narratives before they achieve widespread social saturation [1, 2]. The empirical identification of a median transition latency of approximately two weeks for an emerging weak signal to mature into a dominant narrative defines a critical, finite window of opportunity during which authorities can deploy highly effective prebunking, debunking, and targeted clarification campaigns [14]. This temporal biomarker approach is especially vital given the increasing rhetorical sophistication of modern vaccine-skeptical discourse, which has mutated from localized clinical concerns into complex pseudo-scientific, genetic, and geopolitical conspiracy theories, requiring dynamic, context-aware

policy responses rather than static communication models. Concurrently, because platform architectures facilitate the organized, infrastructure-driven propagation of these themes, governance interventions should shift from individual content moderation to the detection and disruption of coordinated sharing infrastructures [19]. Ultimately, establishing robust digital health surveillance requires integrating these automated early-warning systems with multimodal data pipelines (such as video and voice content) and hosting findings in FAIR-compliant institutional repositories to ensure long-term scientific transparency, reproducibility, and ethical data stewardship [20].

6 LIMITATIONS

This study provides a large-scale quantitative analysis of vaccine discourse on Telegram, but several limitations should be acknowledged. The analysis was restricted to approximately 330,000 text messages, excluding a substantial volume of image, video, and audio content where conspiratorial narratives often emerge. The dataset consists solely of public Telegram messages, limiting generalizability to other platforms and communication environments. Methodologically, BERTrend relies on independent weekly BERTopic models, which may not ensure perfect semantic alignment over time, while results remain sensitive to UMAP and HDBSCAN parameter settings, particularly during periods of low activity. The definition of emerging trends can vary across domains. Findings are specific to the Brazilian Portuguese context and the 2020–2025 period, requiring further validation before being extended to other languages, cultures, or platforms.

7 CONCLUSION

This study demonstrates the effectiveness of the proposed framework for monitoring emerging misinformation narratives, showing that 69.2% of weak signals successfully consolidated in out-of-sample temporal testing and that positive weekly growth trends increased the likelihood of consolidation by 43.46 times. The results reveal the evolution of vaccine-related discourse toward more sophisticated pseudo-scientific and conspiracy-based narratives and identify a median two-week intervention window for public health responses. Future work will focus on benchmarking against classical methods, validating cross-platform generalizability, and integrating advanced capabilities such as embedding-based topic tracking, LLM-assisted classification, automated fact-checking, real-time monitoring, and knowledge graph analysis.

8 ETHICAL CONSIDERATIONS

This study adhered to ethical and privacy standards by analyzing only publicly available data from Unicamp's REDU repository [4] and restricting data collection to public Telegram channels and groups, thereby excluding private and encrypted conversations. According to the dataset authors, user and channel identifiers were pseudonymized using persistent, non-reversible SHA-256 hashes with a private salt. In addition, personally identifiable information (PII), including phone numbers, email addresses, and physical addresses, was automatically redacted using Microsoft Presidio. To further mitigate the risk of re-identification, all timestamped join and leave events were removed from the dataset.

ACKNOWLEDGMENTS

Generative AI tools were used for manuscript revision. The authors acknowledge the original dataset contributors and the Unicamp Research Data Repository (REDU), from which the data were obtained [4]. The dataset documentation is available for non-commercial academic use <https://doi.org/10.25824/redu/5JIVDT>, in accordance with applicable ethical and privacy requirements.

REFERENCES

- [1] Ali F Al-Qahtani, Hanaa Abbas, Stefano Cresci, and Roberto Di Pietro. 2025. The COVID-19 Infodemic: A Survey. *IEEE Access* (2025).
- [2] Abdulrahman Alharbi, Rafaa Aljurbua, Shelly Gupta, Hussain Otudi, Jovan Andjelkovic, and Zoran Obradovic. 2025. Cross-platform spatiotemporal sentiment trends analysis of COVID-19 vaccine discourse. *Social Network Analysis and Mining* 15, 1 (2025), 104.
- [3] Allaa Boutaleb, Jerome Picault, and Guillaume Grosjean. 2024. BERTrend: neural topic modeling for emerging trends detection. In *Proceedings of the Workshop on the Future of Event Detection (FuturED)*. 1–17.
- [4] João Phillipe Cardenuto, Ana Carolina Pontalti Monari, Michelle Diniz Lopes, Leopoldo Andre Dutra Lusquino Filho, and Anderson de Rezende Rocha. 2025. Brazilian Social Media Anti-vaccine Information Disorder Dataset - Telegram. doi:10.25824/redu/5JIVDT
- [5] Athus Cavalini and Giovanni Comarela. 2025. Um Estudo sobre Espalhamento de Desinformação em Comunidades Antivacina no Telegram. In *Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (CTD 2025)*. SBC – Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil. VII Concurso de Teses e Dissertações.
- [6] Huarong Chen, Yadong Wu, Jing Lei, Weixin Zhao, Guijuan Wang, Jing Liao, and Fupan Wang. 2025. Dynamic Topic Analysis and Visual Analytics for Trajectory Data: A Spatial Embedding Approach. *Electronics* 14, 24 (2025), 4873.
- [7] Kelly Simone Almeida Cunegundes, Daisy Maria Machado, and Nadia Vitorino Vieira. 2025. “Não sou antivacina, mas...”: entendendo a hesitação vacinal a partir das narrativas de pais hesitantes. *Cadernos de Saúde Pública* 41 (2025), e00154624.
- [8] Jônatas H. dos Santos, Julio C. S. Reis, Philipe de F. Melo, João F. H. Olivetti, Thales H. Silva, Matheus Guimaraes, Glaucio de Souza, Marco A. G. Rodrigues, Marcos A. Goncalves, Fabricio Benevenuto, and Cristiano X. Lima. 2025. Caracterização da (Des)Informação das Narrativas sobre Vacinação em Grupos Públicos no WhatsApp no Brasil. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia 2025)*. SBC – Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil.
- [9] Juliana da Costa Feitosa, Luiz Felipe de Camargo, Eloisa Bonatti, Giovanna Simioni, and José Remo Ferreira Brega. 2023. eXplainable Artificial Intelligence in Sentiment Analysis of Posts about Covid-19 Vaccination on Twitter. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia '23)*. ACM, New York, NY, USA. doi:10.1145/3617023.3617033
- [10] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure (arXiv: 2203.05794). arXiv.
- [11] Nimesh Gyawali, Doina Caragea, Cornelia Caragea, and Saif M Mohammad. 2025. The shifting landscape of vaccine discourse: Insights from a decade of pre-to post-COVID-19 vaccine posts on social media. *Plos one* 20, 12 (2025), e0337911.
- [12] Sinchul Jwa, Yousuke Imanishi, Marie T Ascher, and Matthew Z Dudley. 2025. Communication interventions to reduce parental vaccine hesitancy: A systematic review. *Vaccine* 61 (2025), 127401.
- [13] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).
- [14] Giorgio Martalò, Marco Menale, and Romina Travaglini. 2026. Controlling misinformation spread through debunking and pre-bunking strategies. *Ricerche di Matematica* (2026), 1–16.
- [15] André Medeiros Mendes and Luiz Gomes-Jr. 2022. Analysis of Misinformation in the Context of the Coronavirus Vaccination in Brazil. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia '22)*. ACM, New York, NY, USA. doi:10.1145/3539637.3558042
- [16] Denise Salotti Augusto Pizani, Márcio Antônio Francisco Dearo, and Aline Ferreira de Oliveira Pereira. 2026. Queda da cobertura vacinal no Brasil: causas, consequências e estratégias de enfrentamento. *Arquivos de Asma, Alergia e Imunologia* 9, 2 (2026), 216–226.
- [17] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
- [18] Julie Ricard, Ergon Cugler de Moraes Silva, Ivette Yañez, and Letícia Hora. 2025. From Theory to Telegram: Enabling Factors of Anti-vaccine Disinformation in Brazil. In *Conference on Digital Government Research*, Vol. 26.
- [19] Yunya Song, Yin Zhang, Sheng Zou, Xian Yang, and Qintao Huang. 2025. The spread of pro-and anti-vaccine views by coordinated communities on facebook during COVID-19 pandemic. *Journal of Computational Social Science* 8, 4 (2025), 98.
- [20] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan Willem Boiten, Luiz Bonino Da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship: Comment. *Scientific data* 3 (2016), 160018.